# 高波の極値統計解析に対する*L*-moments 法の適用と分布関数の選択について

Application of *L*-moments Method to Statistical Analysis of Extreme Wave Heights and Discussion on Selection of Distribution Functions

合田良実[1]・久高将信[2]

Yoshimi GODA and Masanobu Kudaka

[1]土木学会名誉会員 工博 横浜国立大学名誉教授 (株) エコー（〒110-0014 東京都台東区北上野 2-6-4）
[2]土木学会会員 (株) エコー（〒110-0014 東京都台東区北上野 2-6-4）

Application of the L-moments method for statistical analysis of extreme wave heights extracted by the POT method is made for storm wave data at eight stations around Japan, which have the record length of 27 to 38 years. The formulas of the L-moments of the Weibull distribution are newly derived for estimation of the shape, scale and location parameters from a sample. A new index called TUD (Twenty-Up Deviation) is developed for judgment of the degree of goodness of data fitting to a candidate distribution. Among eight stations, two stations show the best fitting to the Weibull distribution, four to the General Pareto (GPA) distribution, and two to the General Extreme Value (GEV) distribution. At two stations fitted to the GPA distribution, the theoretical upper bounds of extreme wave heights are only 7% to 11% greater than the observed maximum wave height.

***Key Words:*** *Extreme wave statistics, GEV, GPA, Weibull distribution, L-moments, NOWPHAS, TUD,*

## １．まえがき

極値統計解析にはいろいろな手法が開発されてきており，極値データの標本への当てはめ方法についてもProbability Weighted Moments (PWM，確率重み付き積率法) がGreenwoodほか (1979) によって提案され，米国における河川流量解析等に使用されてきた。さらに，Hosking (1990) はこのPWM法を発展させた*L*-moments法を発表している。米国においては河川流量の解析では単一河川のみを扱うのではなく，特性をほぼ同じくする流域内の複数の河川が類似の極値分布関数に従うと見なして解析する手法が主流であり，その際には*L*-moments法を用いた研究が多い。地域ごとに共通の極値分布関数を当てはめて解析する方式は地域頻度解析 (Regional Frequency Analysis) と名付けられており，広域の海岸ごとに高波の母分布関数を見いだそうとする合田 (2008, 374頁) のアプローチと類似の目的を持っている。

*L*-moments法を用いた地域頻度解析についてはこれを解説した書籍をHosking・Wallis (1997) が出版しており，また数理統計プログラムのデータベースであるhttp://lib.stat.cmu.edu の中のlmomentsから計算を実行するためのFORTRANプログラムをダウンロードできるようになっている。水文分野以外では外山・水野 (2002) が全国アメダス地点の確率降水量の解析に用いており，高波についてはvan Gelderほか (2000) がオランダ沿岸の波浪データをこの手法を使って解析している。

本論文では*L*-moments法を用いた極値統計解析の方法について解析し，NOWPHASのわが国沿岸の波浪データベースから抽出した高波の資料について適用した事例を紹介する。ただし，Hosking・Wallis (1997) が推奨している*L*-moments法には，高波の極値統計で多用さ

れてきたワイブル分布が取り上げられていない。これまでの各港湾における高波の極値解析資料との連続性からみて，*L*-moments 法を適用するに当たってもワイブル分布を含める必要がある。そこで，本論文ではワイブル分布に対する *L*-moments の理論値を示し，その適用方法を紹介する。

## ２．*L*-moments の定義とその計算

*L*-moments というのは，順序統計量の線形結合 (linear combinations of order statistics) として Hosking (1990) が名付けたものである。いま，確率変量 $X$ の分布関数を $F(X)$ とし，$X$ がある閾値を超えない確率が $P$ であるような閾値を $x(P)$ と表示する。$X(P)$ は $X$ の quantile function と呼ばれ，波高の極値統計でいうところの再現確率統計量である。これを数式で表示すれば

$$F[x(P)] = P, \quad x = F^{-1}(P) \tag{1}$$

この確率統計量すなわち quantile function を用いて，まず次の確率重み付き積率 $\beta_r$ を定義する。

$$\beta_r = \int_0^1 x(P)\, P^r dP \tag{2}$$

*L*-moments はこの積率 $\beta_r$ を用いて次式のように定義される。

$$\lambda_1 = \beta_0 \tag{3}$$

$$\lambda_2 = 2\beta_1 - \beta_0 \tag{4}$$

$$\lambda_3 = 6\beta_2 - 6\beta_1 + \beta_0 \tag{5}$$

$$\lambda_4 = 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \tag{6}$$

さらに，上記の *L*-moments から相互の比率を次の用に定義する。

*L*-CV (Coefficient of *L*-variation)：　$\tau = \lambda_2 / \lambda_1$ (7)

*L*-skewness (*L*-ひずみ度)：　$\tau_3 = \lambda_3 / \lambda_2$ (8)

*L*-kurtosis (*L*-尖鋭度)：　$\tau_4 = \lambda_4 / \lambda_2$ (9)

1 次の *L*-moment $\lambda_1$ は母集団の平均値を与え，2 次の *L*-moment $\lambda_2$ は標準偏差に関するものの偏差そのものではなく，母集団の標準偏差を $\sigma$ で表すと $0.5\sigma \le \lambda_2 \le \sigma/3^{0.5}$ の範囲の値をとる (Hosking · Wallis (1997), p. 35 より推定)。

式 (2)～(9) は母集団に関するものである。いま，高波の観測・推算資料からからある閾値を超える極値時系列データ (POT) が $n$ 個，すなわち大きさが $n$ の標本が得られたとすると，これを昇順に並べ替えて順序統計量に変換する。すなわち，$x_1 \le x_2 \le \ldots \le x_n$ と並べ替える。この順序統計量から次のような $r$ 次平均量 $b_r$ を計算する。

$$b_0 = \frac{1}{n}\sum_{j=1}^{n} x_j \tag{10}$$

$$b_1 = \frac{1}{n}\sum_{j=2}^{n} \frac{(j-1)}{(n-1)} x_j \tag{11}$$

$$b_2 = \frac{1}{n}\sum_{j=3}^{n} \frac{(j-1)(j-2)}{(n-1)(n-2)} x_j \tag{12}$$

一般に，

$$b_r = \frac{1}{n}\sum_{j=3}^{n} \frac{(j-1)(j-2)\ldots(j-r)}{(n-1)(n-2)\ldots(n-r)} x_j \tag{13}$$

参考として，lmoments のデータファイルの中の年最大風速の一つの標本（フロリダ州タンパ）について計算した結果を**表−1** に示す。左から第 2 欄が各年の最大風速，第 3 欄が順序統計量に並べ替えたものであり，第 4～6 欄が $b_r$ を計算するための行である。$b_r$ の計算結果は最下行に示されている。

**表−1**　$n = 10$ の場合の $r$ 次平均量 $b_r$ の計算表

| $j$ | $x$ | $b_0 = x_j$ | $b_1$ | $b_2$ | $b_3$ |
|---|---|---|---|---|---|
| 1 | 45 | 37 | | | |
| 2 | 47 | 42 | 4.666667 | | |
| 3 | 65 | 44 | 9.777778 | 1.222222 | |
| 4 | 50 | 45 | 15.00000 | 3.750000 | 0.535714 |
| 5 | 56 | 47 | 20.88889 | 7.833333 | 2.238095 |
| 6 | 55 | 50 | 27.77778 | 13.88889 | 5.952381 |
| 7 | 37 | 53 | 35.33333 | 22.08333 | 12.61905 |
| 8 | 53 | 55 | 42.77778 | 32.08333 | 22.91667 |
| 9 | 44 | 56 | 49.77778 | 43.55556 | 37.33333 |
| 10 | 42 | 65 | 65.00000 | 65.00000 | 65.00000 |
| | | 49.4 | 27.1 | 18.94167 | 14.65952 |

この $r$ 次平均量 $b_r$ を積率 $\beta_r$ の普遍推定値であり，標本 *L*-moments の値は，$b_r$ を用いて次のように計算する。

$$\ell_1 = b_0 \tag{14}$$

$$\ell_2 = 2b_1 - b_0 \tag{15}$$

$$\ell_3 = 6b_2 - 6b_1 + b_0 \tag{16}$$

$$\ell_4 = 20b_3 - 30b_2 + 12b_1 - b_0 \tag{17}$$

そして，この標本 *L*-moments を用いて次のような標本 *L*-moments 比を算定する。

*L*-CV (Coefficient of L-variation)：　$t = \ell_2 / \ell_1$ (18)

*L*-skewness (*L*-ひずみ度)：　$t_3 = \ell_3 / \ell_2$ (19)

*L*-kurtosis (*L*-尖鋭度)：　$t_4 = \ell_4 / \ell_2$ (20)

**表−1** の例では，標本 *L*-moments と標本 *L*-moments 比が

次のように算定される。

$\ell_1 = 49.4,\quad \ell_2 = 4.8,\quad \ell_3 = 0.450,\quad \ell_4 = 0.7405,$

$t = 0.0972,\quad t_3 = 0.0938,\quad t_4 = 0.1543$

なお，標本に対して $l_1, l_2, \ldots, t, t_3, t_4, \ldots$ などの記号を用いるのは，母分布関数に対する推定値であることを明示するためである。

## ３．極値分布関数と *L*-moments 比の関係

Greenwood ほか (1979) による PWM 法は，標本について確率を重みとした積率を計算することによって適合候補の分布関数の形状・尺度・位置母数を速やかに推定することを目的としていた。*L*-moment 法においても，各種の極値分布関数について *L*-moments の理論値を計算し，分布関数の母数との関係を求めておくことによって分布関数の母数推定が可能となる。

Hosking (1990) は式 (2)～(9) を用いて各種の極値分布関数に対する $\lambda_1$, $\lambda_2$, $\tau_3$, および $\tau_4$ の理論値ならびに母数の推定式を計算した。この結果は Hosking・Wallis (1997) の解説書にも付録としてまとめられている。そのうちから，高波の極値統計解析に使用する分布関数に関わるものを以下に再録する。ただし，ワイブル分布については Greenwood ほか (1979) を参照して今回新しく導いたものである。

(A) 指数型分布

分布関数：

$$F(x) = 1 - \exp\{-(x-B)/A\},\quad B \le x < \infty \tag{21}$$

確率統計量 (quantile)：

$$x = F^{-1}(P) = B - A\ln(1-P) \tag{22}$$

*L*-moments：

$$\lambda_1 = B + A,\ \lambda_2 = (1/2)A,\ \tau_3 = 1/3,\ \tau_4 = 1/6 \tag{23}$$

母数推定式：

$$A = 2\lambda_2,\quad B = \lambda_1 - A \tag{24}$$

(B) ガンベル分布（極値 I 型分布）

分布関数：

$$F(x) = \exp[-\exp\{-(x-B)/A\}],\ B \le x < \infty \tag{25}$$

確率統計量 (quantile)：

$$x = F^{-1}(P) = B - A\log(1-\ln P) \tag{26}$$

*L*-moments：

$$\lambda_1 = B + A\gamma, \lambda_2 = A\ln 2,\ \tau_3 \cong 0.1699,\ \tau_4 \cong 0.1504 \tag{27}$$

母数推定式：

$$A = \lambda_2/\ln 2,\quad B = \lambda_1 - A\gamma \tag{28}$$

ここに $\gamma$ はオイラーの定数, 0.5772....である。

(C) 一般化極値分布 (Generalized Extreme-Value Distribution: GEV)

分布関数：

$$F(x) = \begin{cases} \exp[-\{1-k(x-B)/A\}^{1/k}], & k>0:\ -\infty \le x < B + A/k \\ & k<0:\ B - A/k < x < \infty \\ \exp[-\exp\{-(x-B)/A\}], & k=0:\ -\infty < x < \infty \end{cases} \tag{29}$$

確率統計量 (quantile)：

$$x = F^{-1}(P) = \begin{cases} B + A\{1-(-\ln P)^k\}/k, & k \ne 0 \\ B - A\ln(-\ln P), & k=0 \end{cases} \tag{30}$$

*L*-moments：

$$\begin{aligned} &\lambda_1 = B + A\{1-\Gamma(1+k)\}/k,\quad \lambda_2 = A(1-2^{-k})\Gamma(1+k)/k, \\ &\tau_3 = 2(1-3^{-k})/(1-2^{-k}) - 3, \\ &\tau_4 = \{5(1-4^{-k}) - 10(1-3^{-k}) + 6(1-2^{-k})\}/(1-2^{-k}) \end{aligned} \tag{31}$$

母数推定式：

$$\begin{aligned} &k \approx 7.8590c + 2.9554c^2,\quad c = \frac{2}{3+\tau_3} - \frac{\log 2}{\log 3} \\ &A = \frac{\lambda_2 k}{(1-2^{-k})\Gamma(1+k)},\quad B = \lambda_1 - A\{1-\Gamma(1+k)\}/k \end{aligned} \tag{32}$$

一般化極値分布は，式 (29) の表現では $k < 0$ のときに極値 II 型分布，$k > 0$ のときに極値 III 型分布となり，前者では確率統計量が下限値を持ち，後者では上限値を持つ。また，$k = 0$ のときは極値 I 型分布である。ただし，高波の極値統計で使われている極値 II 型は式 (29) の形状母数を $k' = -1/k$ と変換したものに等しい。

なお，式 (32) における形状母数 $k$ は近似推定式であり，Hosking (1990) が導いたものである。

(D) 一般化パレート分布 (Generalized Pareto Distribution: GPA)

分布関数：

$$F(x) = \begin{cases} 1-[1-k(x-B)/A]^{1/k} & 1 \ge k > 0:\ B \le x < B + A/k \\ & k<0:\quad B \le x < \infty \\ 1-\exp\{-(x-\xi)/\alpha\}, & k=0:\quad B \le x < \infty \end{cases} \tag{33}$$

確率統計量 (quantile)：

$$x = F^{-1}(P) = \begin{cases} B + A\{1-(1-P)^k\}/k, & k \ne 0 \\ B - A\ln(1-P), & k=0 \end{cases} \tag{34}$$

*L*-moments：

$$\begin{aligned} &\lambda_1 = B + A/(1+k),\quad \lambda_2 = A/\{(1+k)(2+k)\}, \\ &\tau_3 = (1-k)/(3+k),\quad \tau_4 = (1-k)(2-k)/\{(3+k)(4+k)\} \end{aligned} \tag{35}$$

母数推定式：

$$k=(1-3\tau_3)/(1+\tau_3),\quad A=(1+k)(2+k)\lambda_2,\quad B=\lambda_1-(2+k)\lambda_2 \tag{36}$$

一般化パレート分布は一般化極値分布のような極大・極小値に対する漸近関数ではないけれども，ポアソン分布に従う確率変量のうち，ある閾値を超えるものに着目する場合などを表示するのに適切な分布関数といわれる。北野ほか (2002) はKodiak沖の高波データについてこの分布の適用性を検討している。なお，この分布関数の特性等についてはColes (2001) などを参照されたい。

(E) ワイブル分布

分布関数：

$$F(x)=1-\exp[-\{(x-B)/A\}^k],\quad B<x<\infty \tag{37}$$

確率統計量(quantile)：

$$x=F^{-1}(P)=B+A[\{-\log(1-P)\}^{1/k}] \tag{38}$$

$L$-moments：

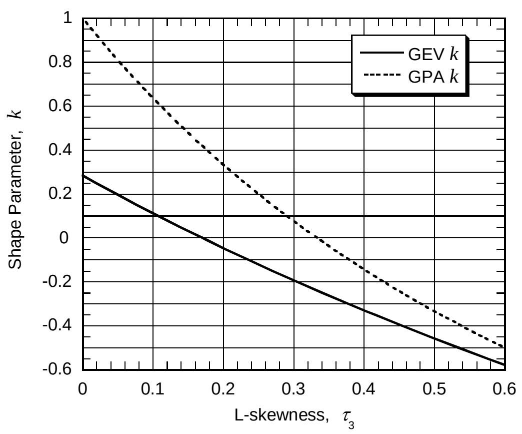
$$\lambda_1=B+A\Gamma(1+1/k),\quad \lambda_2=A(1-2^{-1/k})\Gamma(1+1/k),$$
$$\tau_3=3-2(1-3^{-1/k})/(1-2^{-1/k}),$$
$$\tau_4=\{5(1-4^{-1/k})-10(1-3^{-1/k})+6(1-2^{-1/k})\}/(1-2^{-1/k}) \tag{39}$$

母数推定式：

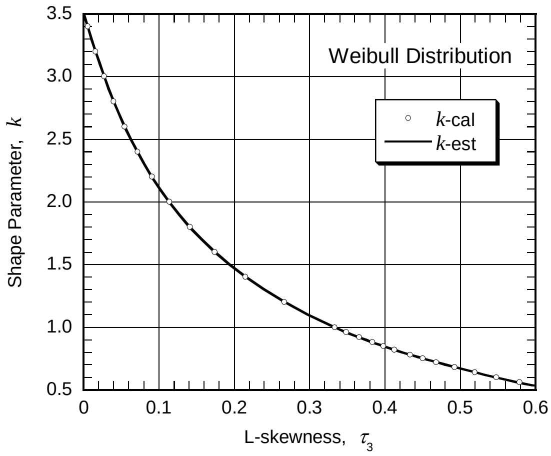
$$k=285.3\,\lambda_3^6-658.6\,\lambda_3^5+622.8\,\lambda_3^4-317.2\,\lambda_3^3+98.52\,\lambda_3^2-21.256\,\lambda_3+3.5160 \tag{40}$$

$$A=\frac{\lambda_2}{(1-2^{-1/k})\Gamma(1+1/k)},\quad B=\lambda_1-A\Gamma(1+1/k) \tag{41}$$

極値分布関数の特性を決めるのは形状母数$k$である。$L$-moments法では$L$-skewnessの値によって母数$k$の値が一義的に定まる。**図−1**はGPAとGEVについて母数$k$と$L$-skewnessの関係を示したものである。




**図−1**　GPAおよびGEV分布における形状母数$k$と$L$-skewness $\tau_3$との関係




**図−2**　ワイブル分布における形状母数$k$と$L$-skewness $\tau_3$との関係

ワイブル分布の形状母数$k$を推定する式(40)は著者らが数値計算結果から導いた近似式である。すなわち，形状母数$k$をいろいろ変えて3次の積率$\lambda_3$を計算し，その結果に対して6次の多項式を当てはめたものである。近似精度は±0.2%である。

**図−2**はワイブル分布における母数$k$と$L$-skewnessの関係を示している。なお，この図では数値計算の結果と式(40)による推定値の両方を示しているが，図上では両者の違いを見分けることがむずかしい。

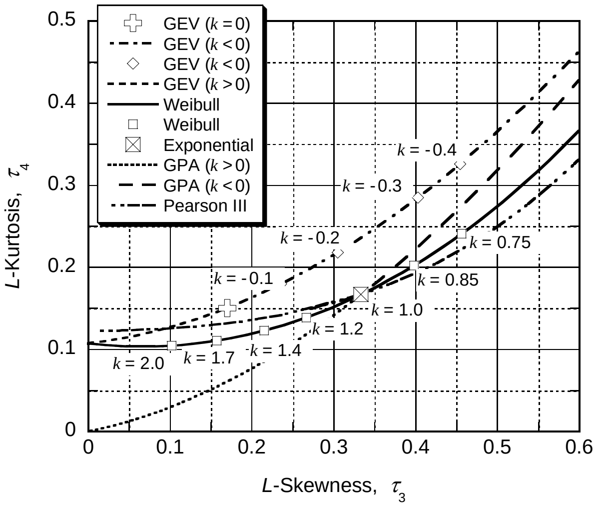
## 4．$L$-moments比に基づく極値分布関数の選別

前節に紹介した極値分布関数のうち，指数分布と極値I型分布は2母数型分布であって，$L$-skewnessと$L$-kurtosisは一定値である。他の分布関数は3母数型分布であるため，形状母数$k$の値によって$L$-skewnessと$L$-kurtosisの値が連続的に変化し，両者の関係は分布関数に特有な一つの曲線で表される。ここでは紹介していないが，4母数型分布分の場合には，両者がある面内で変化する関係を示す。

**図−3**は，前節の分布関数に加えてPearson III型分布も含め，二つの$L$-skewness比の関係をプロットしたものである。2母数型分布の指数型分布とガンベル分布はそれぞれ一つの点で表される。指数型分布の点（四角にクロス）から左下へ延びる曲線は一般化パレート分布(GPA)で形状母数が$k>0$の場合を表し，同じく右上へ上昇する曲線は$k<0$の場合である。なお，$k>0$の場合には前述のように確率統計量$x$の値に上限がある。式(33)の関数形はHosking・Wallis (1997)によるもので，一般の数理統計学では形状母数の符号が逆であるのが普通である。

ガンベル分布(FT-I)の点（白十字）の左へ緩やかに

下がる線は一般化極値分布 (GEV) のうちで確率変数 $x$ の値に上限がある場合に相当し，極値III型分布 (FT-III) である。右へ緩やかに上がる線は極値II型分布 (FT-II) である。ワイブル分布は極値III型分布と上限値がある一般化パレート分布の間に位置するが，ワイブル分布は上限値がなく，右裾がどこまでも延びるのが特徴的である。




**図－3** 各種分布関数における*L*-skewnessと*L*-kurtosisの関係

図中のワイブル分布と極値II型分布については，形状母数に特定の値を与えたときの*L*-skewnessと*L*-kurtosisの対応点を示してある。また，ワイブル分布とPearson III型分布は*L*-moments比の値が隣接している。すなわち，後者はワイブル分布で代替させることが可能であり，極値統計の標本当てはめの候補に入れなくても良いと考えられる。

ワイブル分布と一般化パレート分布 (GPA) は指数分布を共有している。指数分布の点の左下側では，両者の*L*-kurtosisの値に差異が見られる。すなわち，ワイブル分布では*L*-kurtosisの値が余り変化しないのに対し，一般化パレート分布では*L*-skewnessの減少につれて*L*-kurtosisも減少する。指数分布の点の右上側ではワイブル分布と一般化パレート分布の違いが小さいが，後者のほうが*L*-kurtosisがやや大きめの値を示す。

このように，*L*-skewnessと*L*-kurtosisの二つのパラメータを用いることによって，極値分布関数の差異を判別できる可能性が考えられる。**図－3**に示したのは母集団についての結果であり，データ数に限りのある標本の場合には統計的変動性の影響によって母集団ほどの差別化はむずかしいであろう。

地域頻度解析の場合には，複数地点の極値資料を収集し，これらが同じ母集団に属することを検定した上で，複数地点全体に共通な分布関数を見いだす方式を用いている。すなわち，それぞれの標本に含まれているデータ数の少なさを複数地点全体のデータ数で補う方式である。

地域頻度解析は年最大値資料を対象としており，高波のように極値時系列法 (POT) で抽出されたデータに対してどのように地域頻度解析を行うかは今後の課題である。本論文ではまず単一地点の極値資料に対してあらかじめ分布関数の候補を選定し，*L*-momentsを用いて形状・尺度・位置の3母数を推定する方式について検討を進める。以下においては，この方式をLMN法と略称する。

## 5．使用する高波資料の概要

先に筆頭著者はNOWPHASの1989年までのデータベースを使用し，日本沿岸の高波の極値分布関数について考察した（合田 2008, pp. 374-375）。それから既に10年を経過し，波浪データベースも拡充されたところから，長期間にわたる波浪データが蓄積されている8地点を選び，POT法で抽出した高波資料の解析を行った。これらの高波資料の概要を**表－1**に示す。

表－1　検討に用いた高波資料の概要

| 地点名 | 観測期間 | 年数 | データ数 | 波高閾値 (m) | $(H_{1/3})_{max}$ (m) |
|---|---|---|---|---|---|
| 酒田 | 1970-2008 | 38 | 923 | 3.0 (0.5) 5.0 | 10.65 |
| 金沢 | 1970-2008 | 37 | 722 | 3.0 (0.5) 5.0 | 8.14 |
| むつ小川原 | 1971-2008 | 37 | 551 | 2.5 (0.5) 5.0 | 9.56 |
| 鹿島 | 1972-2008 | 36 | 598 | 2.5 (0.5) 5.0 | 7.50 |
| 波浮 | 1973-2008 | 35 | 640 | 3.0 (0.5) 5.0 | 8.48 |
| 志布志 | 1980-2008 | 28 | 482 | 1.5 (0.5) 4.0 | 10.30 |
| 中城湾 | 1975-2008 | 27 | 599 | 2.0 (0.5) 5.0 | 12.08 |
| 那覇 | 1973-2008 | 35 | 656 | 2.5 (0.5) 4.5 | 9.24 |

観測年数は1年単位の欠測期間を除外したものであるが，月単位の欠測期間の調整は行っていない。また，合田・小長谷 (1998) が行ったような，高波のピーク時の部分欠測も補填していない。また，気象原因別（台風系と低気圧系など）のデータ区分もおこなっていない。したがって，この資料だけでは確率波高の推定作業を進めることはできない。

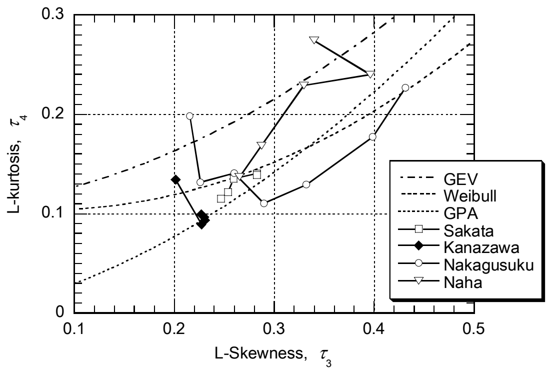
表中のデータ数は，波高閾値が最も低いときのものであり，それから0.5 mごとに閾値を高めていったので，解析を行ったデータ数は閾値ごとに異なる。

## 6．LMN法における最適合分布の判定

極値統計解析では，与えられた標本に対して複数の分布関数を当てはめ，そのうちで適合度の最も高い関数を

仮に母集団分布と見なす方式が一般的である。最小2乗法の場合には，確率変量とその基準化変量との間の相関係数が高いほど適合度が良いとみなす（これを若干修正したのが合田が提案したMIR基準である）。最尤法では尤度を判定基準に使うけれども，最適関数判定の感度がそれほど良いものではない。

LMN法の場合には，まず*L*-skewnessの値によって分布関数ごとの形状母数が算定される。そうすると，その分布関数の*L*-kurtosisの理論値が計算されるので，その値と標本から求めた値との差を判定基準とする方式が一つ考えられる。**図−5**は，沿岸8地点のうちの4地点について波高閾値を変えることによって，*L*-skewnessと*L*-kurtosisの関係がどのように変わるかも示したものである。

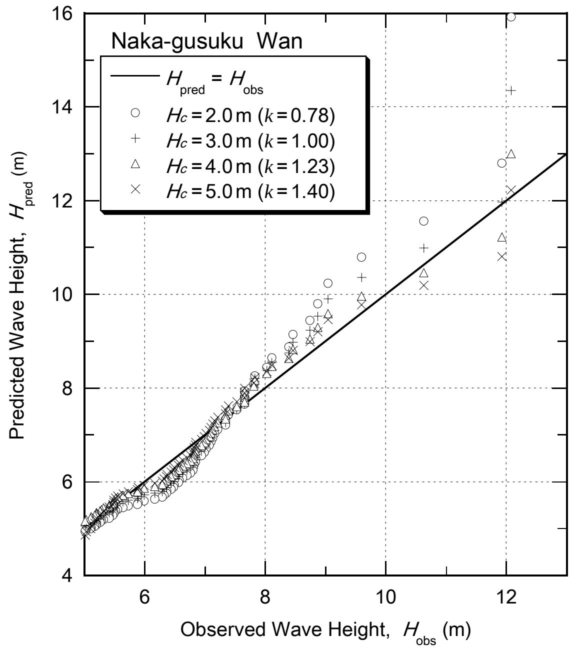



**図−5**　波高閾値を変えたときの沿岸波浪の*L*-skewnessと*L*-kurtosisの関係

酒田港では$H_c = 3.0$ mのデータが左下に位置し，閾値を上げるにつれて表示点がワイブル分布の曲線に沿って右上に移動する。金沢港では$H_c = 5.0$ mのデータが一つだけ左上に離れており，他の閾値のデータは表示点がGPAの曲線上の一カ所に固まっている。中城湾の資料では，$H_c = 2.0$ mのデータが右上に位置し，閾値を上げるにつれて表示点がワイブル分布の曲線の下方を左下へ向けて大きく移動し，$H_c$=5.0 mのときにはGEVの関係曲線の左上へ飛び跳ねる。那覇港の場合には，$H_c = 3.0$ mのデータが左下に位置し，閾値を上げるにつれて表示点がGEVの曲線へ向かって大きく右上に移っていく。

波高の閾値によって適合分布関数が変化することはあり得るけれども，**図−5**に示す中城湾や那覇港のケースは変化が極めて大きく，単一の極値資料に対して*L*-skewnessと*L*-kurtosisの関係だけに基づいて最適合の分布関数を選ぶ方式には疑問を抱かせる。実際に**表−1**の高波資料について，標本から得られた*L*-kurtosisの値とその理論値との差を分布関数ごとに比較し，最低値を示す分布関数を調べてみると，波高の閾値を上げるたびに異なる最適合分布が得られてしまう。

たとえば，**図−6**は中城湾の高波データについて波高の閾値を $H_c = 2.0$ (1.0) 5.0 mと4通りに変えたときの元の波高データと，確率統計量として推定された波高値とを比較した図である。データは波高の大きさの順に並べられておりこうした表示法はQuantile – Quantile (Q – Q)プロットと呼ばれる。ここでは当てはめ分布関数をワイブル分布に固定しているが，波高閾値によって形状母数が変化している。




**図−6**　中城湾のデータに波高閾値を変えてワイブル分布を当てはめたときの順序統計量と確率統計量の比較 (Q–Qプロット)

LMN法では，式 (1)〜(4) の$r$次平均量$b_r$を計算する際に，データを小さい方から昇順に並べ替えた順序統計量を使う。閾値を上げたときにはその閾値よりも小さな波高データは切り捨てられ，それよりも大きなデータの順位が切り下げられる。$n$個のデータの中で小さい方から$j$番目，あるいは大きい方から$m$番目のデータに対する確率統計量を求めるときには，その非超過確率$P_{j:n}$を与える必要があり，このために次のプロッティング・ポジション公式を使う。

$$P_{j:n} = \frac{j-0.35}{n}, \quad P_{m:n} = 1 - \frac{m-0.65}{n} \qquad (42)$$

式 (42) の非超過確率 $P_{j:n}$ は GPA 分布に対して Hosking・Wallis (1997) が推奨 (33頁) しているもので，他の分布関数については問題があるかもしれないと述べている。しかし，他に適切な公式が提示されていないので，本論文ではQ – Qプロットを行うときの非超過確率

として式 (42) を使うことにする。なお，$P_{m:n}$ は $n = j + m - 1$ の関係から導かれたものである。

最小2乗法の場合には，プロッティング・ポジションの取り方によって確率統計量に bias が生じることがあるため，公式の選定には十分な注意が必要であるけれども，LMN 法では分布関数の当てはめ自体にはプロッティング・ポジション公式の選択は影響を与えない。

**図−6**では，あらかじめ分布関数としてワイブル分布を固定した条件で，極値波高データを抽出する閾値の影響を調べたものである。図から明らかなように，閾値の設定によって確率統計量が観測された波高の順序統計量と一致する度合いが相当に変化する。

設計波高を選定するために極値統計解析を行う場合には，主として波高の上位のデータに対する極値分布関数の適合度に関心が持たれる。そこで，まず極値データの観測値の順序統計量と確率統計量の相対偏差 (%) を次のように定義する。

$$E_m = \left( \frac{H_{pred,m} - H_{obs,m}}{H_{obs,m}} \right) \times 100 \tag{43}$$

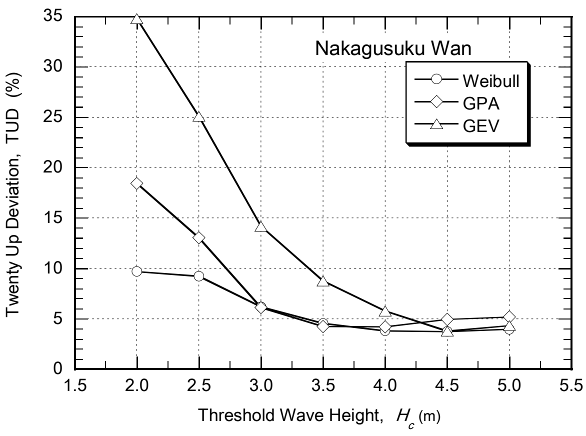
ここに，$H_{\mathrm{pred},\,m}$ は上から第 $m$ 番目の確率統計量，$H_{\mathrm{obs},\,m}$ は観測波高の第 $m$ 番目の順序統計量である。そして，上位20個の2乗平均偏差を次のように定義し，それが最小となるものを最適合とすることを提案する。

$$TUD = \sqrt{\frac{1}{20}\sum_{m=1}^{20} E_m^2} \tag{44}$$

記号の TUD は，Twenty-Up Deviation の頭文字を取った略号である。20個を取り上げたことに特別の理論的根拠はない。各種分布関数を当てはめたときの適合度を比較し，最もよく適合するものを選ぶ判断基準として使うためのものである。上位10個についても試行してみたところ，TUD の絶対値はやや大きくなったものの，分布関数ごとの適合度の順位は変わらなかった。

中城湾の高波極値データについて TUD による適合度を計算すると**図−7**のようになり，高波抽出の閾値を上げるに従って適合度が高まることが分かった。**図−7**では，当てはめ候補の極値分布関数として3母数型分布である GVE，GPA, およびワイブル分布の3種類を取り上げている。**3**節では5種類の関数を紹介したけれども，そのうちの指数型分布は GPA 分布およびワイブル分布に含まれているので，上記の3分布関数を用いれば良いと考えたものである。先に**図−5**に示した中城湾の *L*-skewness と *L*-kurtosis の関係の変化を**図−7**の TUD の値の変化から読みとることはむずかしい。

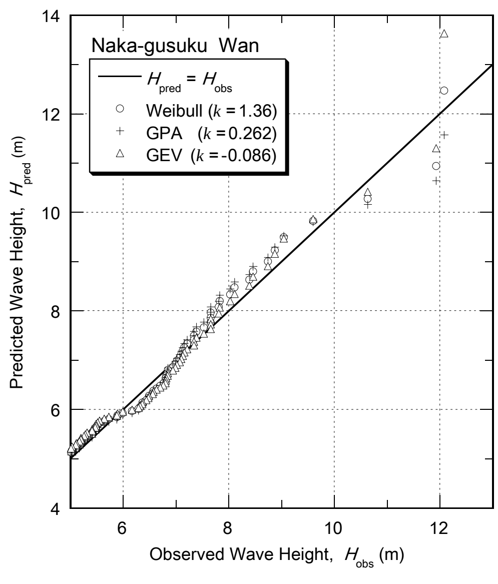
高波抽出の閾値によって適合度が変化するのは，中城湾の場合には台風系と低気圧系という母集団の異なる高波が混在しているためである。合田・小長谷 (1978) の場合には抽出された高波の全てについて気象原因を確認し，台風系とそれ以外の高波に区分して解析したけれども，今回はそうした吟味を行っていない。波高の閾値が低い場合には波高が比較的に小さい非台風系のデータが数多く含まれ，閾値を上げるに従って台風系の高波のみが解析対象となったものであろう。**図−7**によれば，閾値を $H_c \geq 4.0$ m とすることによって，TUD の値が安定するので，そうしたデータを対象として中城湾の高波の確率波高を計算すればよいと思われる。




**図−7**　中城湾の高波データに対するTUD値

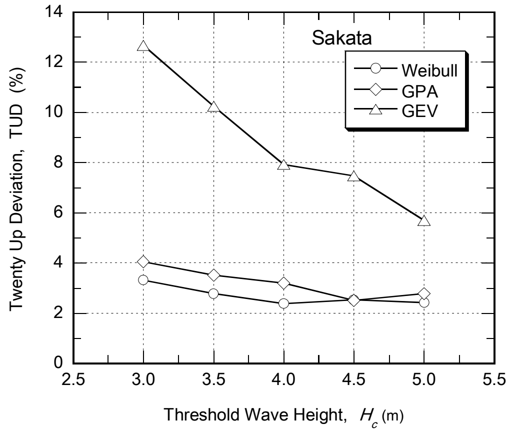
## ７．各地点の高波の極値分布特性

本節では，**表−1**に示した8地点の高波資料に対して LMN 法で極値分布関数を当てはめた結果について述べ

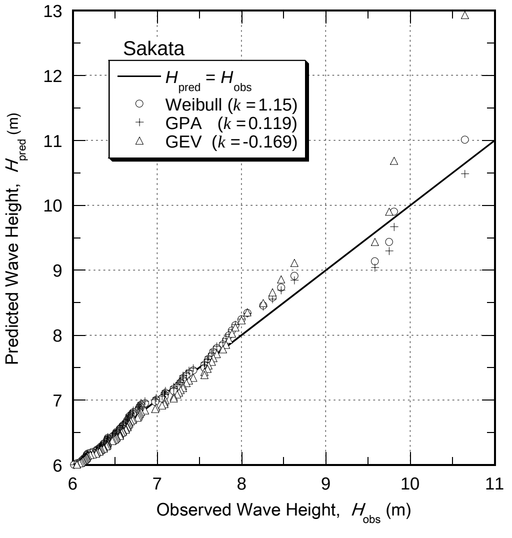



図−8　中城湾の高波 ($H_c$=4.5m) に対する極値分布関数の当てはめ結果

る。**図－8**は中城湾の高波データであって，**図－7**の結果を参照して高波抽出の閾値を$H_c = 4.5$ mと設定した場合を示している。この高波データでは観測値の第1位が12.08 m，第2位が11.93 mと近接しており，このため確率統計量の推定結果が前者に対しては過大，後者に対しては過小な結果となっている。　3種類の分布関数は適合度に大差がなく，どの関数も適用可能である。GEV分布は順序統計量の第1位に対してやや過大な値を与えるけれども，第2位以下に対しては他の二つの分布関数よりも適合度がよい。



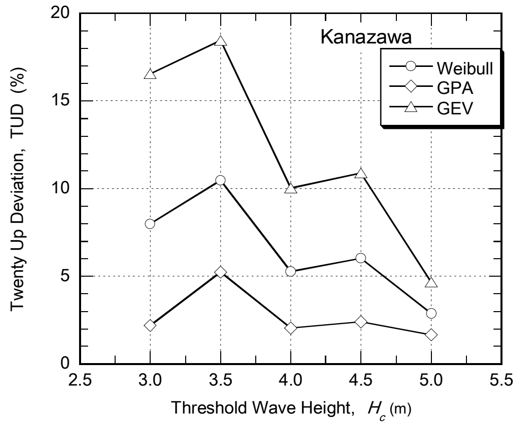

**図－9**　酒田港の高波データに対するTUD値



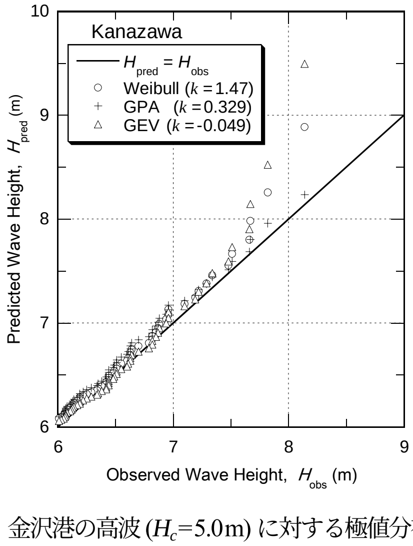

**図－10**　酒田港の高波 ($H_c$=5.0m) に対する極値分布関数の当てはめ結果

次に，酒田港の高波データに対する極値分布関数の適合度指標のTUD値を**図－9**，閾値を$H_c = 5.0$ mと設定した場合の順序統計量と確率統計量の比較を**図－10**に示す。酒田港の場合にはワイブル分布が最適合であり，閾値を変えても適合度はそれほど変化しない。すなわち，酒田港の波浪データはそのほとんどが同じ高波の母集団に所属すると推測される。GPA分布はワイブル分布よりもわずかながら適合度に劣るけれども，$H_c = 5.0$ mの場合の**図－10**では両者の違いはわずかである。GEV分布は上位の順序統計量に対して過大評価，中位の順序統計量に対して過小評価となり，酒田港には不向きである。

金沢港については，極値分布関数の適合度指標のTUD値を**図－11**，閾値を$H_c$ = 5.0 mと設定した場合の順序統計量と確率統計量の比較を**図－12**に示す。




**図－11**　金沢港の高波データに対するTUD値

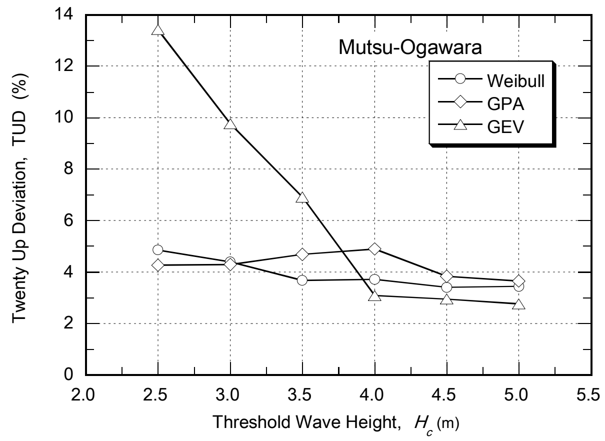



**図－12**　金沢港の高波 ($H_c$=5.0m) に対する極値分布関数の当てはめ結果

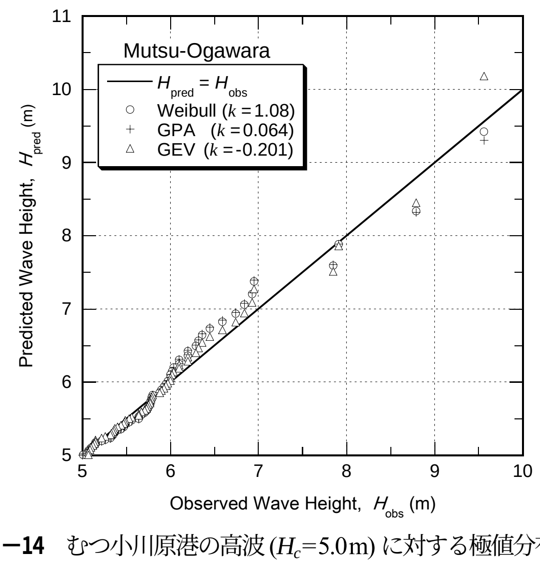
金沢港の高波資料では，閾値を変えてもGPA分布の優位性が変わらない。ワイブル分布やGEV分布を当てはめると，上位の順序統計量に対して過大評価となってしまう。なお，高波抽出の波高閾値が$H_c = 3.5$ mのときに適合度指標のTUD値が増大する理由は解明できていない。ただし，閾値レベルによる適合度の変化が少ないところから，金沢港の高波資料も同じ母集団に属するとみなされよう。

次にむつ小川原港についての検討結果を**図－13，14**に示す。**図－13**は適合度指標のTUD値が波高閾値によっ

て変化する状況を示している。ワイブル分布とGPA分布は波高閾値の影響をほとんど受けず，どちらかといえばワイブル分布のほうが適合度がよい。一方，GEV分布は波高閾値が低いときには適合度が極めて低いけれども，波高閾値が $H_c$ = 4.0 m以上ではワイブル分布やGPA分布よりも高波のデータによく適合するようになる。**図−14**は波高閾値が $H_c$ = 5.0 mのときの順序統計量と確率統計量の比較を示したものである。第1位のデータについてはGEV分布がやや高めの推定値を与えるものの，波高が $H_{obs}$ = 6.0〜7.0 mの範囲ではGEV分布がデータによく適合しており，TUD値が最小となることを裏付けている。



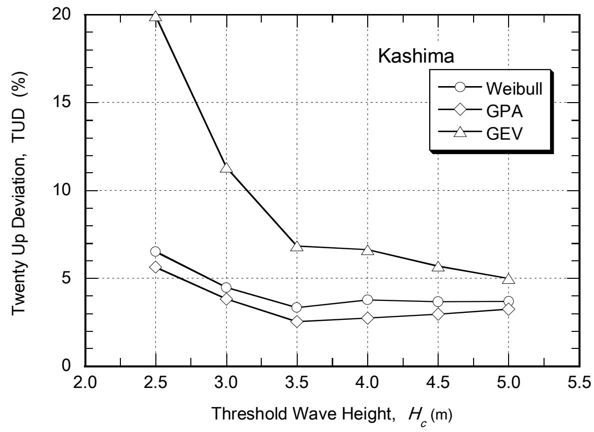

**図−13** むつ小川原港の高波データに対するTUD値



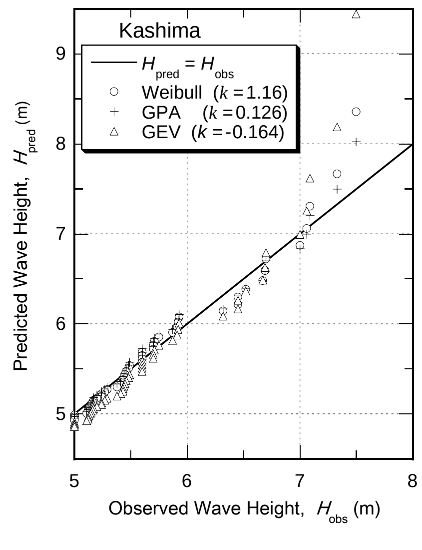

**図−14** むつ小川原港の高波 ($H_c$=5.0m) に対する極値分布関数の当てはめ結果

鹿島港についての検討結果は，**図−15, 16**に示す通りである。**図−15**は適合度指標のTUD値が波高閾値によって変化する状況を示しており，ワイブル分布とGPA分布は波高閾値の影響をあまり受けず，どちらかといえばGPA分布のほうが適合度がよい。一方，GEV分布は全般的に適合度が低く，鹿島港の高波に対する極値分布としては不適当である。他の地点では波高閾値を上げるに従ってTUD値が小さくなる，すなわち波高データに当てはめる分布関数の適合度が高まるのであるが，鹿島港の場合には波高閾値が $H_c$ = 3.5 mのときにTUD値が最小となっている。

この波高閾値が $H_c$ = 3.5 mのときの順序統計量と確率統計量の比較を示したのが**図−16**である。図で明らかなように，GPA分布が波高の観測データからの乖離が最も少なく，他の分布よりもTUD値が小さいことに対応している。




**図−15** 鹿島港の高波データに対するTUD値

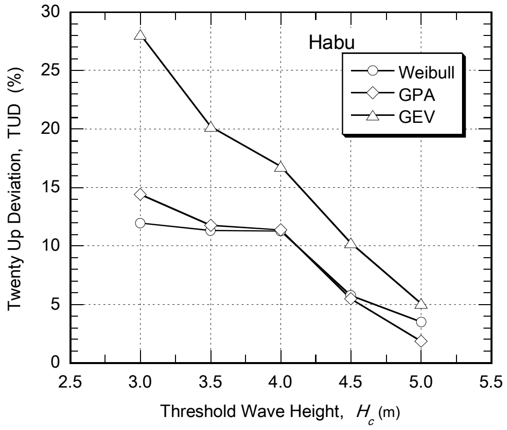



**図−16** 鹿島港の高波 ($H_c$=3.5m) に対する極値分布関数の当てはめ結果

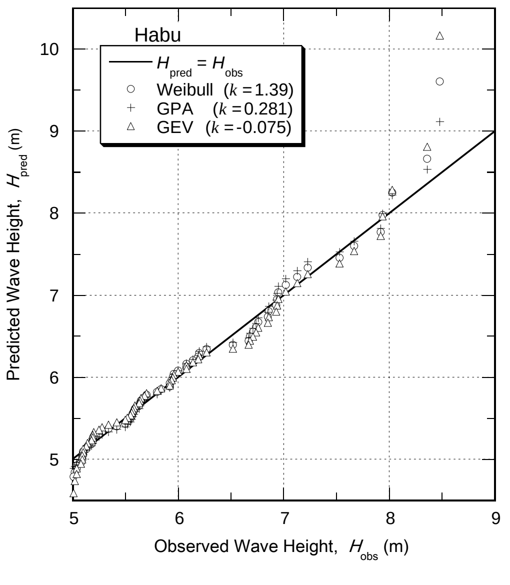
その次は波浮港であり，このデータに対する検討結果は**図−17, 18**に示す通りである。**図−17**は適合度指標のTUD値が波高閾値によって変化する状況を示しており，閾値レベルが上るにつれてTUD値が急速に減少している。中城湾の場合と同様に，波高レベルの低い高波のグ

ループと高いグループとで，高波を発生させた気象原因が異なる可能性が強い。極値分布関数としてGEV分布は全閾値レベルで適合度が低い。ワイブル分布とGPA分布とでは，波高閾値が低いときは前者の適合度が高く，波高閾値が高まると後者のほうが適合度が高くなる。

**図−18**は波高閾値が$H_c = 5.0$mのときの順序統計量と確率統計量の比較を示したものである。GPA分布が高波のデータによく適合していることが認められる。



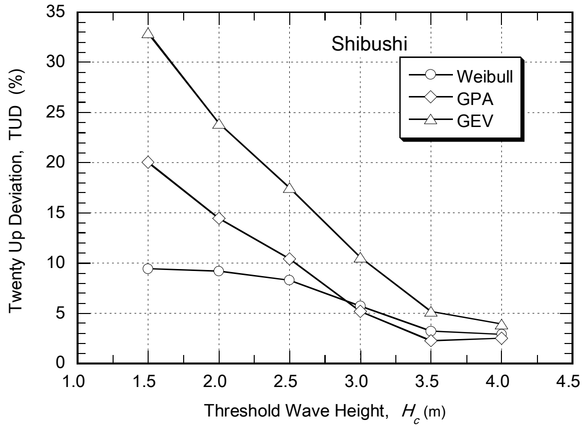

**図−17**　波浮港の高波データに対するTUD値



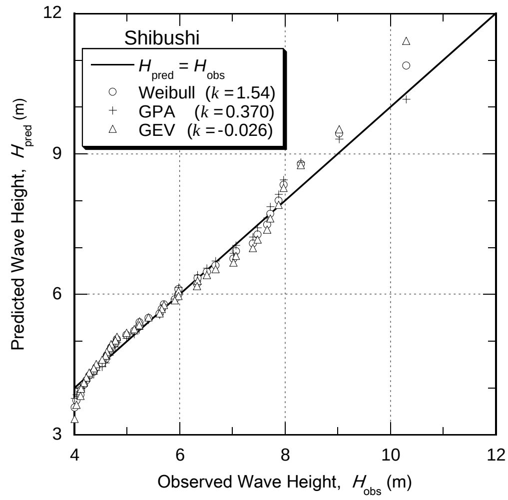

**図−18**　波浮港の高波($H_c$=5.0m)に対する極値分布関数の当てはめ結果

さらに**図−19, 20**には志布志港の高波データに対する検討結果であり，**図−19**は適合度指標のTUD値の変化，**図−20**は波高閾値が$H_c = 4.0$mのときの順序統計量と確率統計量の比較を示している。

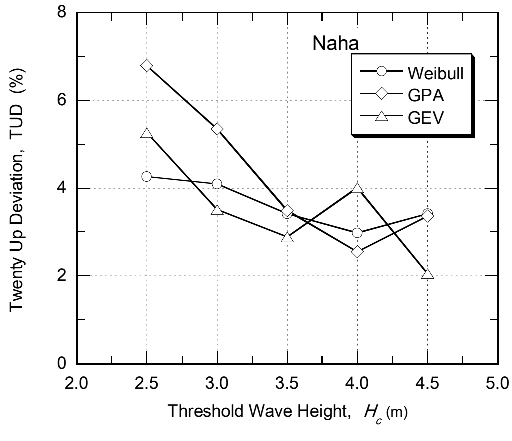
波高閾値による極値分布関数の適合度の変化は波浮港のデータと類似しており，閾値レベルがあがるにつれて適合度が一様に向上する。3種類の分布関数の中ではGEVの適合度が劣り，波高閾値が低い間はワイブル分布が優位であり，波高閾値が高くなるとGPA分布がやや優位に立つ。




**図−19**　志布志港の高波データに対するTUD値

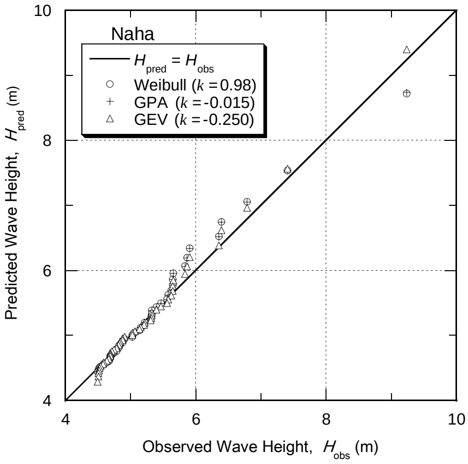



**図−20**　志布志港の高波($H_c$=4.0m)に対する極値分布関数の当てはめ結果




**図−21**　那覇港の高波データに対するTUD値

最後に，那覇港の高波データについての検討結果を**図−21, 22**に示す。**図−21**は適合度指標のTUD値の変化




**図−22**　那覇港の高波 ($H_c$=4.5m) に対する極値分布関数の当てはめ結果

であり，他の諸港とは異なってGEV分布の適合度がワイブル分布やGPA分布よりも高いケースが多く，特に波高閾値が$H_c$ =4.5mのときには適合度が最も高くなっている。この状況は，**図−22**の順序統計量と確率統計量の比較によく現れている。

## 8．GPA 分布における上限値の存在について

今回の検討結果では，一般化パレート (GPA) 分布が適合する地点がかなり多かった。波高閾値の選定にもよるが，適合度指標のTUD値が最小であったのが8地点では4地点を数えた。しかし，GPA分布は式 (33) の表式で形状母数が正の時に極値が上限値$x_{upper} = B + A/k$で頭打ちになるという特性がある。そこで，この理論的上限値が観測最大波高に対してどの程度離れているかを調べたのが**表−2**である。

**表−2**　GPA分布を適用したときの波高の上限値

| 地点 | $H_c$ (m) | $N$ | $k$ | $A$ (m) | $B$ (m) | $H_{upper}$ (m) | $(H_{obs})_{max}$ (m) |
|---|---|---|---|---|---|---|---|
| 酒田 | 5.0 | 244 | 0.1187 | 1.212 | 4.97 | 15.18 | 10.65 |
| 金沢 | 5.0 | 162 | 0.3294 | 1.236 | 4.98 | 8.73 | 8.14 |
| むつ小川原 | 5.0 | 63 | 0.0644 | 0.977 | 4.99 | 20.15 | 9.56 |
| 鹿島 | 3.5 | 214 | 0.1265 | 1.110 | 3.47 | 12.24 | 7.50 |
| 波浮 | 5.0 | 66 | 0.2811 | 1.547 | 4.87 | 10.38 | 8.48 |
| 志布志 | 4.0 | 42 | 0.3704 | 2.868 | 3.74 | 11.48 | 10.30 |
| 中城湾 | 4.5 | 88 | 0.2621 | 2.440 | 4.45 | 13.76 | 12.08 |
| 那覇 | 4.5 | 51 | –0.015 | 0.8331 | 4.42 | ∞ | 9.24 |

**表−2**の地点のうち，那覇港のデータに対しては形状母数が負値であるので上限値はない。また，酒田港と中城湾についてはワイブル分布が最適合，むつ小川原港と那覇港についてはGEV分布が最適合である。その他の4地点のうち，金沢港では波高が$H_{upper}$ = 8.73 mで頭打ちになると予測され，既往最大観測値の8.14 mとの差が僅か (7%) しかない。また，志布志港でも $(H_{obs})_{max}$ = 10.30 mに対して$H_{upper}$ = 11.48 mとその差が小さい (11%)。また，両地点における高波の発生特性を考えると，発生波高がこれらの上限値で抑えられるということは考えにくい。

極値分布関数の当てはめの適合度だけで判定すると一般化パレート (GPA) 分布が最適合となっても，設計のための確率波高を選定する際には，上限値を持たないワイブル分布を当てはめることを優先すべきであると考えられる。この点については今後さらに検討したい。

## 9．地域頻度解析のための今後の課題について

今回の検討結果では，地点ごとに最適合と判定される極値分布関数が異なり，またPOT法で高波データを抽出するときの閾値によっても最適合の分布関数が逆転することも少なくなかった。日本沿岸の設計波高を合理的に設定するためには，高波の母集団についての検討をさらに高度なものとする必要がある。

最初に述べたように，米国の河川流量解析では母集団が同じと推定される複数の河川の流量データを総合的に解析する地域頻度解析の手法が用いられている。この方法は毎年最大値を取り扱っており，各河川の流量データをそれぞれの平均値で除して正規化した無次元流量を対象としてLMN法を適用する。複数河川の無次元流量のデータに対して共通の極値分布関数が得られたならば，河川ごとの平均流量を乗じることによって流量の極値分布を復元する。

これに対して高波の極値統計解析では，データ数を多くして確率統計量の信頼区間を狭めるためにPOT法を用いるのが標準であり，単純に地点ごとの高波データの平均値で正規化することには疑問が残る。POT法にふさわしい地域頻度解析法を開発する必要がある。

また，今回の検討では波高の閾値レベルを上げることによって分布関数の適合度が向上したけれども，それに付随して利用できるデータの個数が減少している。今回は信頼区間について検討していないけれども，信頼区間の幅はデータ個数の平方根に逆比例するので，データはできるだけ多く残してデータ個数を多くしておきたい。分布関数の適合度と信頼区間の幅の兼ね合いをどのように取るのかが今後の課題である。

Hosking・Wallis(1997)が提唱している地域頻度解析では，最適分布関数が定まった後でモンテカルロ法の試行によって信頼区間幅を推定するようになっている。彼ら

の方式では，特定の地域内の複数の標本に適合する分布関数の候補として4母数あるいは5母数型含めているので，モンテカルロ法以外に信頼区間幅を推定することは困難であろう。しかし，地域共通の分布関数の候補を3母数型に限定するならば，あらかじめモンテカルロ法を駆使して信頼区間幅を推定するための経験式を導くことも可能であろう。

いずれにしても，まずは高波のデータを気象原因別に分類することから作業を進める必要があろう。

## １０．まとめ

本論文は，高波の極値統計解析に *L*-moments 法を適用するための一つの試論であり，今後検討を進めるべき課題が多く残されている。ここまでの段階で明らかになった事項を列挙すると以下のようになろう。

1) *L*-moments 法でこれまで考慮されていなかったワイブル分布を当てはめ候補の極値分布関数にふくめることができた。
2) 標本から計算される *L*-skewness に基づいてワイブル分布，一般化極値 (GEV) 分布，および一般化パレート (GPA) 分布の形状・尺度・位置母数の3母数を求める方法が明確化された。
3) 標本に対する極値分布関数の適合度判定の指標として，順序統計量の上位20個の2乗平均偏差を用いるTUD (Twenty-Up Deviation) を新たに考案した。
4) 日本沿岸で長期間にわたる波浪観測データが利用できる8地点について *L*-moments 法による極値統計解析を試みたところ，波高の閾値レベルによるTUD変化から高波の母集団が単一あるいは複数混在を判別する可能性が示された。
5) 8地点中でワイブル分布を最適とするものが2地点，GPA 分布を最適とするものが4地点，GEV 分布を最適とするものが2地点であった。
6) GPA 分布は適合する地点が多いものの，確率統計量が上限値で頭打ちとなる危険性が例証された。

## 謝　辞

本論文で使用した高波のデータは港湾空港技術研究所で所管しているNOWPHASのデータベースに基づくものであり，データ使用を承認してくださった河合弘泰海象情報研究領域長のご厚意に深謝する次第である。

## 参考文献

北野利一・間瀬　肇・喜岡　渉・矢野陽一郎 (2002)：一般化パレート分布による極値波浪解析－拡張形状母数の推定－，海岸工学論文集，第49巻，　pp. 161-165.

合田良実 (2008)：「耐波工学－港湾・海岸構造物の耐波設計－」，鹿島出版会，430p.

合田良実・小長谷　修・永井紀彦 (1998)：極値波浪統計の母分布関数に関する実証的研究，海岸工学論文集，第45巻，pp.211-215.

外山奈央子・水野　量 (2002)：*L*-moments を用いた地域頻度解析による全国アメダス地点における確率降水量の推定，気象庁研究時報，54巻 5－6号合併号，pp. 55-100.

Coles, S. (2001): *An Introduction to Statistical Modeling of Extreme Values*, Springer, 208p.

Greenwood, J A., Landwehr, J. M., Matalas, N. C., and J. R. Wallis (1978): Probability weighted moments: Definition and relation to parameters of several distributions expressable in inverse form, *Water Resources Res.*, Vol. 15, No. 5, pp. 1049-1064.

Hosking, J. R. M. (1990): *L*-moments: Analysis and estimation of distributions using linear combinations of order statistics, *J. Roy. Statistical Soc.*, *Series B*, **52**, pp. 105-24.

Hosking, J. R. M. and J. R. Wallis (1997): "*Regional Frequency Analysis*," Cambridge Univ. Press, 224p.

van Gelder, P. H. A. J. M. (2000): *Statistical Methods for the Risk-Based Design of Civil Structures*, Ph.D Thesis Delft University of Technology, 249p.

van Gelder, P. H. A. J. M, J. De Ronde, and N. W. Neykov (2000): Regional frequency analysis of extreme wave heights: trading space for time, *Coastal Engineering 2000* (*Proc. 26th ICCE*, Sydney), ASCE, pp. 1099-1112.